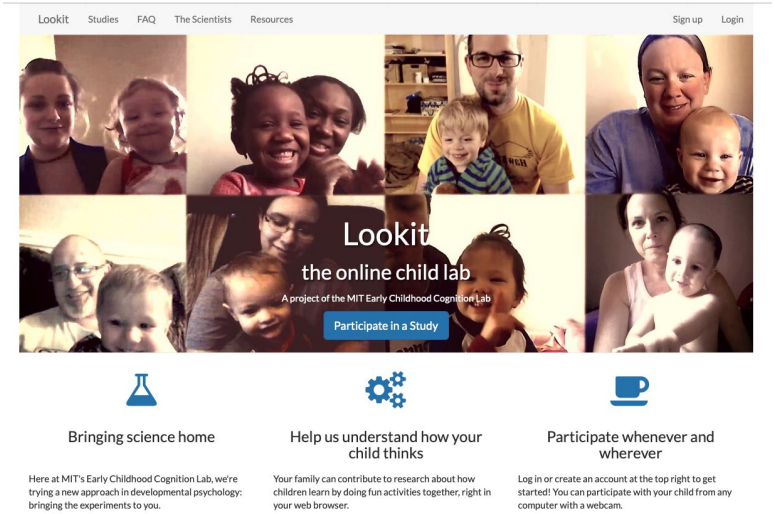
Automated Gaze Coding for Infant Videos

Xincheng Tan, Peng Cao

Motivation

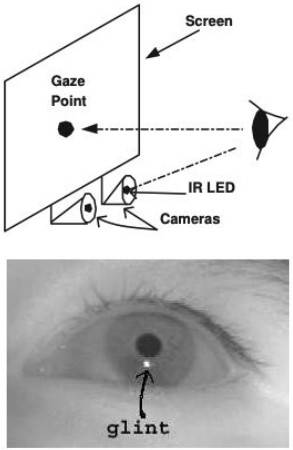
Preferential Looking Paradigm



Lookit Online Platform

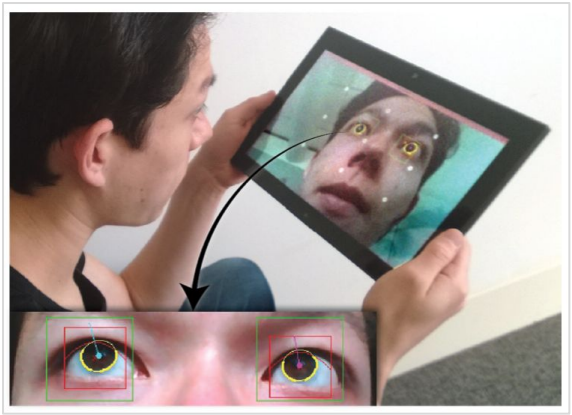
Related work: three categories

Requires additional hardware!



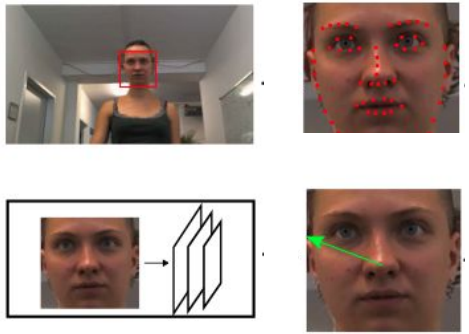
Feature-based
(Zhu *et al.*, 2005)

Requires high video quality!



Model-based
(EyeTab, Wood & Bulling, 2014)

Requires fine-grained annotations!



Appearance-based
(OpenGaze, Zhang *et al.,* 2019)

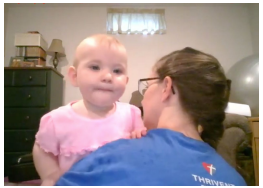
iCatcher (Erel *et al.*, 2020)

Left

Right

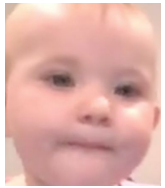
Away

Pipeline:



Video Frame

Face Extractor



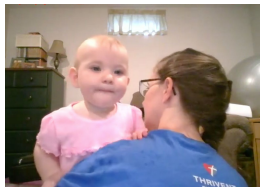
Selected Image Patch

Gaze Estimator

Away / **Left** / Right

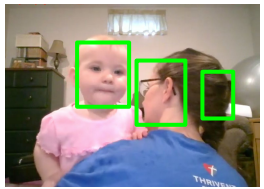
Gaze Direction Label

iCatcher



Video Frame

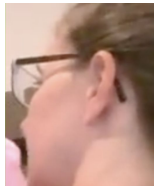
Face Extractor



Extracted Faces

Wrong patch

Lowest Face



Selected Image Patch

Loss of positional information

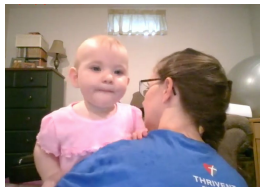
Gaze Estimator

Error!

Away / Left / Right

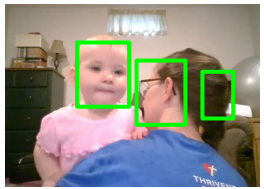
Gaze Direction Label

Proposed Framework



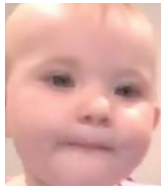
Video Frame

1. Face Extractor

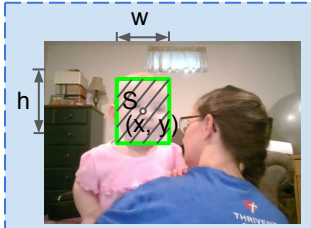


Extracted Faces

2. Infant Selector



Selected Image Patch



Positional Information

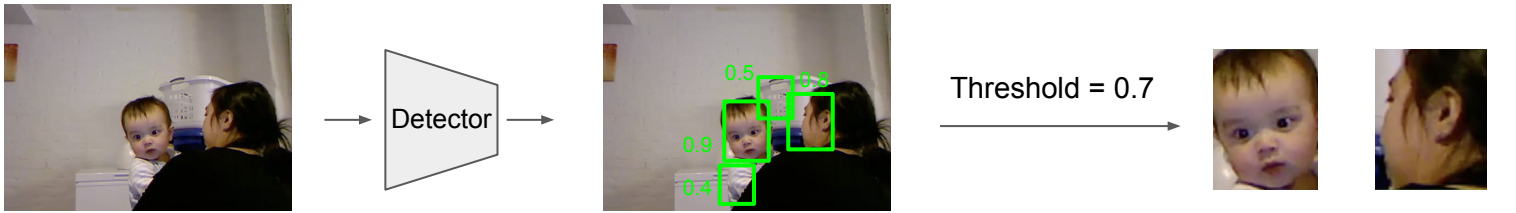
3. Gaze Estimator

Correct!

Away / **Left** / Right

Gaze Direction Label

1. Face Extractor



Video Frame

Face Detection with Confidence Score

Extracted Image Patches

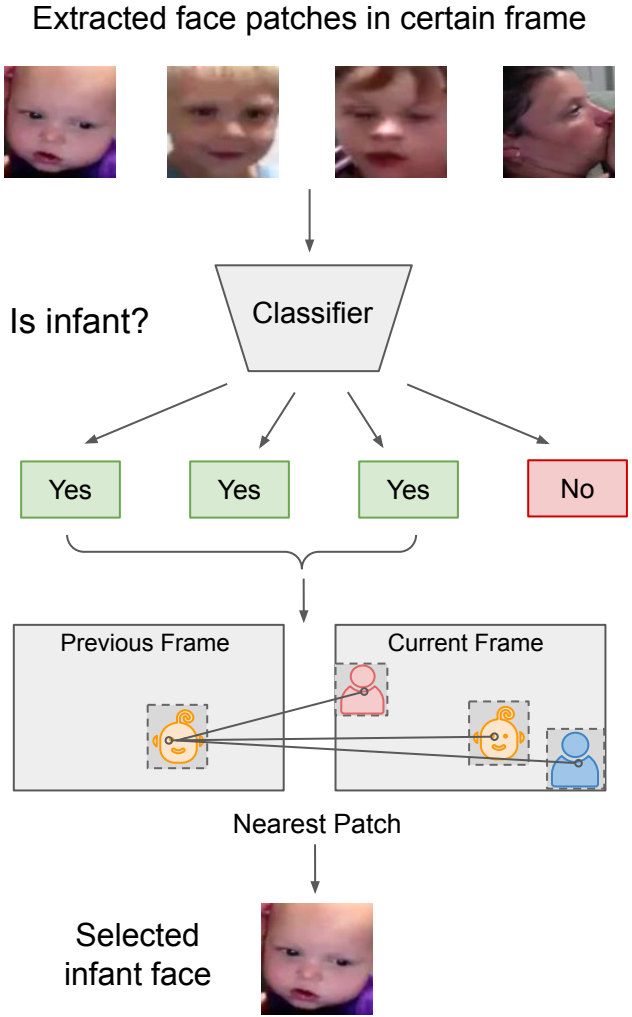
Example image patches from Lookit dataset:

Frame 1 Frame 2 Frame 3 Frame 4 Frame 1 Frame 2 Frame 3 Frame 4

2. Infant Selector

Manually curated dataset:
400 training images, 200 test images, half for each class

Classifier	Accuracy
Lowest-face	60.5%
VGG-11	92.5%
VGG-16	**95.0%**
ResNet-18	94.0%
ResNet-34	91.5%
Wide ResNet	87.5%



Infant Classifier: Failure Cases

Type I Error: undetected infants

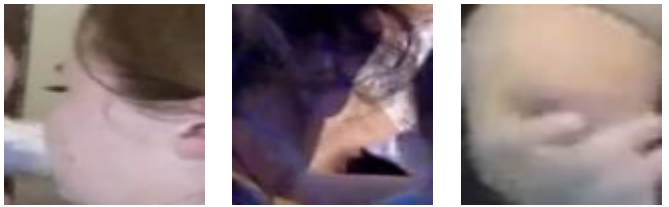
- By default, no-infant frames are assigned “away” label
- Coincides with when infant is looking away

Type II Error: “fake” infants

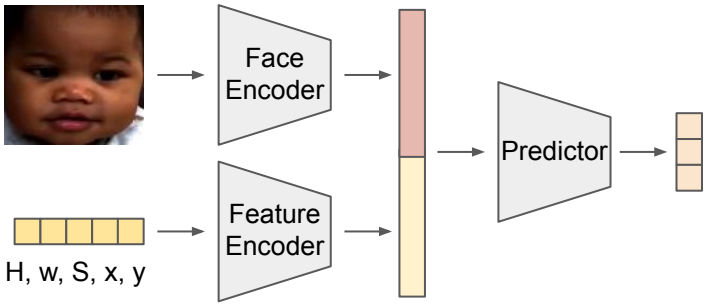
- Multiple infant faces per frame
- Corrected by nearest patch

undetected infants

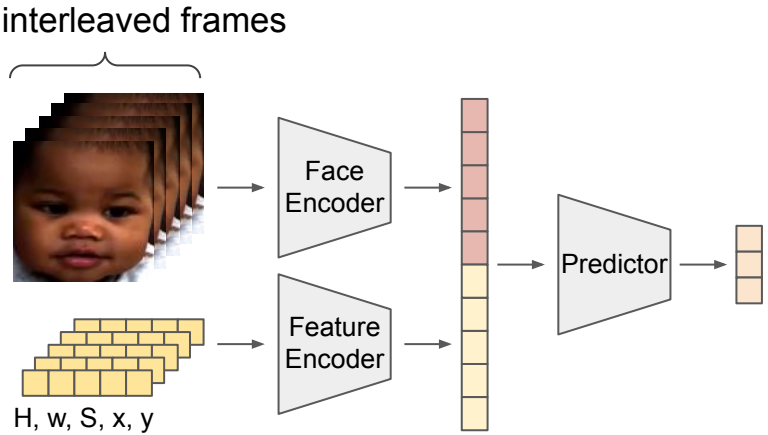
“fake” infants



3. Gaze Estimator



Single-frame gaze estimator



Multi-frame gaze estimator

Face Encoder: ResNet-18
Feature Encoder: 2 fully-connected layers
Predictor: 3 fully-connected layers

Gaze Direction Classification Accuracy

Full Lookit Dataset: Training set: 600,000 frames; Test set: 400,000 frames

	Lowest-face		Ours
Single-frame	82.14%	<	83.58%
Single-frame + Positional Info.	82.23%	<	**84.20%**
Multi-frame	84.25%	<	85.61%
Multi-frames + Positional Info.	84.65%	<	**85.95%**
Multi-frame(E)	86.23%	<	88.11%
Multi-frame(E) + Positional Info.	86.98%	<	**88.58%**

E: Eliminating all the *transition* datapoints (gaze direction classes change within the datapoint).

Limitation and Future Works

Very small dataset for the infant classifier

Transfer Learning

Face extraction suffer from occlusions

Developing a framework without face extraction
e.g. using eye extraction

Transition datapoints

Calibration step:
Ask the infants to look at something that moves around the screen boundary.

Thank you!

Questions?